Chapter 10

Hypothesis Testing: Two Means, Paired Data, Two Proportions

10.1 Hypothesis Testing: Two Population Means and Two Population Proportions[1]

10.1.1 Student Learning Objectives

By the end of this chapter, the student should be able to:

- Classify hypothesis tests by type.
- Conduct and interpret hypothesis tests for two population means, population standard deviations known.
- Conduct and interpret hypothesis tests for two population means, population standard deviations unknown.
- Conduct and interpret hypothesis tests for two population proportions.
- Conduct and interpret hypothesis tests for matched or paired samples.

10.1.2 Introduction

Studies often compare two groups. For example, researchers are interested in the effect aspirin has in preventing heart attacks. Over the last few years, newspapers and magazines have reported about various aspirin studies involving two groups. Typically, one group is given aspirin and the other group is given a placebo. Then, the heart attack rate is studied over several years.

There are other situations that deal with the comparison of two groups. For example, studies compare various diet and exercise programs. Politicians compare the proportion of individuals from different income brackets who might vote for them. Students are interested in whether SAT or GRE preparatory courses really help raise their scores.

In the previous chapter, you learned to conduct hypothesis tests on single means and single proportions. You will expand upon that in this chapter. You will compare two averages or two proportions to each other. The general procedure is still the same, just expanded.

---

[1]This content is available online at <http://http://cnx.org/content/m17029/1.7/>.

To compare two averages or two proportions, you work with two groups. The groups are classified either as **independent** or **matched pairs**. **Independent groups** mean that the two samples taken are independent, that is, sample values selected from one population are not related in any way to sample values selected from the other population. **Matched pairs** consist of two samples that are dependent. The parameter tested using matched pairs is the population mean. The parameters tested using independent groups are either population means or population proportions.

> NOTE: This chapter relies on either a calculator or a computer to calculate the degrees of freedom, the test statistics, and p-values. TI-83+ and TI-84 instructions are included as well as the test statistic formulas. Because of technology, we do not need to separate two population means, independent groups, population variances unknown into large and small sample sizes.

This chapter deals with the following hypothesis tests:

Independent groups (samples are independent)

- Test of two population means.
- Test of two population proportions.

Matched or paired samples (samples are dependent)

- Becomes a test of one population mean.

10.2 Comparing Two Independent Population Means with Unknown Population Standard Deviations[2]

1. The two independent samples are simple random samples from two distinct populations.
2. Both populations are normally distributed with the population means and standard deviations unknown unless the sample sizes are greater than 30. In that case, the populations need not be normally distributed.

The comparison of two population means is very common. A difference between the two samples depends on both the means and the standard deviations. Very different means can occur by chance if there is great variation among the individual samples. In order to account for the variation, we take the difference of the sample means, $\overline{X_1}$ - $\overline{X_2}$, and divide by the standard error (shown below) in order to standardize the difference. The result is a t-score test statistic (shown below).

Because we do not know the population standard deviations, we estimate them using the two sample standard deviations from our independent samples. For the hypothesis test, we calculate the estimated standard deviation, or **standard error**, of **the difference in sample means**, $\overline{X_1}$ - $\overline{X_2}$.

 The standard error is:

$$\sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}$$

(10.1)

The test statistic (t-score) is calculated as follows:

 T-score

$$\frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}}$$

(10.2)

where:

---

[2]This content is available online at <http://http://cnx.org/content/m17025/1.14/>.

- s_1 and s_2, the sample standard deviations, are estimates of σ_1 and σ_2, respectively.
- σ_1 and σ_2 are the unknown population standard deviations.
- $\overline{x_1}$ and $\overline{x_2}$ are the sample means. μ_1 and μ_2 are the population means.

The **degrees of freedom (df)** is a somewhat complicated calculation. However, a computer or calculator calculates it easily. The dfs are not always a whole number. The test statistic calculated above is approximated by the Student-t distribution with dfs as follows:

Degrees of freedom

$$df = \frac{\left[\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}\right]^2}{\frac{1}{n_1-1}\cdot\left[\frac{(s_1)^2}{n_1}\right]^2 + \frac{1}{n_2-1}\cdot\left[\frac{(s_2)^2}{n_2}\right]^2} \tag{10.3}$$

When both sample sizes n_1 and n_2 are five or larger, the Student-t approximation is very good. Notice that the sample variances $s_1{}^2$ and $s_2{}^2$ are not pooled. (If the question comes up, do not pool the variances.)

NOTE: It is not necessary to compute this by hand. A calculator or computer easily computes it.

Example 10.1: Independent groups
The average amount of time boys and girls ages 7 through 11 spend playing sports each day is believed to be the same. An experiment is done, data is collected, resulting in the table below:

	Sample Size	Average Number of Hours Playing Sports Per Day	Sample Standard Deviation
Girls	9	2 hours	$\sqrt{0.75}$
Boys	16	3.2 hours	1.00

Table 10.1

Problem
Is there a difference in the average amount of time boys and girls ages 7 through 11 play sports each day? Test at the 5% level of significance.

Solution
The population standard deviations are not known. Let g be the subscript for girls and b be the subscript for boys. Then, μ_g is the population mean for girls and μ_b is the population mean for boys. This is a test of two **independent groups**, two population **means**.

Random variable: $\overline{X_g} - \overline{X_b}$ = difference in the average amount of time girls and boys play sports each day.

H_o: $\mu_g = \mu_b \ (\mu_g - \mu_b = 0)$

H_a: $\mu_g \neq \mu_b \ (\mu_g - \mu_b \neq 0)$

The words **"the same"** tell you H_o has an "=". Since there are no other words to indicate H_a, then assume **"is different."** This is a two-tailed test.

Distribution for the test: Use t_{df} where df is calculated using the df formula for independent groups, two population means. Using a calculator, df is approximately 18.8462. **Do not pool the variances.**

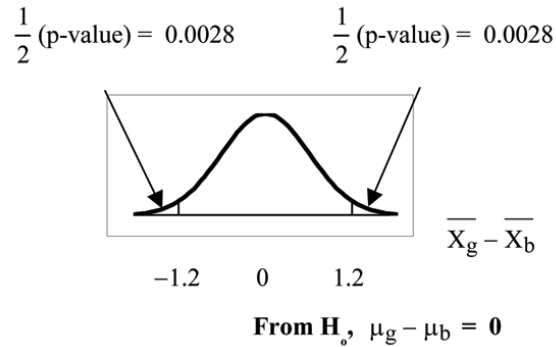**Calculate the p-value using a Student-t distribution:** p-value = 0.0054

Graph:

$\frac{1}{2}$ (p-value) = 0.0028 $\frac{1}{2}$ (p-value) = 0.0028

$\overline{X_g} - \overline{X_b}$

−1.2   0   1.2

From H$_o$, $\mu_g - \mu_b = 0$

Figure 10.1

$s_g = \sqrt{0.75}$

$s_b = 1$

So, $\overline{x_g} - \overline{x_b} = 2 - 3.2 = -1.2$

Half the p-value is below -1.2 and half is above 1.2.

Make a decision: Since $\alpha >$ p-value, reject H_o.

This means you reject $\mu_g = \mu_b$. The means are different.

Conclusion: At the 5% level of significance, the sample data show there is sufficient evidence to conclude that the average number of hours that girls and boys aged 7 through 11 play sports per day is different.

NOTE: TI-83+ and TI-84: Press STAT. Arrow over to TESTS and press 4:2-SampTTest. Arrow over to Stats and press ENTER. Arrow down and enter 2 for the first sample mean, $\sqrt{0.75}$ for Sx1, 9 for n1, 3.2 for the second sample mean, 1 for Sx2, and 16 for n2. Arrow down to μ1: and arrow to does not equal μ2. Press ENTER. Arrow down to Pooled: and No. Press ENTER. Arrow down to Calculate and press ENTER. The p-value is p = 0.0054, the dfs are approximately 18.8462, and the test statistic is -3.14. Do the procedure again but instead of Calculate do Draw.

Example 10.2
A study is done by a community group in two neighboring colleges to determine which one graduates students with more math classes. College A samples 11 graduates. Their average is 4 math classes with a standard deviation of 1.5 math classes. College B samples 9 graduates. Their average is 3.5 math classes with a standard deviation of 1 math class. The community group believes that a student who graduates from college A **has taken more math classes,** on the average. Test at a 1% significance level. Answer the following questions.

Problem 1   *(Solution on p. 460.)*
Is this a test of two means or two proportions?

Problem 2   *(Solution on p. 460.)*
Are the populations standard deviations known or unknown?

Problem 3   *(Solution on p. 460.)*
Which distribution do you use to perform the test?

Problem 4   *(Solution on p. 460.)*
What is the random variable?

Problem 5   *(Solution on p. 460.)*
What are the null and alternate hypothesis?

Problem 6   *(Solution on p. 460.)*
Is this test right, left, or two tailed?

Problem 7   *(Solution on p. 460.)*
What is the p-value?

Problem 8   *(Solution on p. 460.)*
Do you reject or not reject the null hypothesis?

Conclusion:
At the 1% level of significance, from the sample data, there is not sufficient evidence to conclude
that a student who graduates from college A has taken more math classes, on the average, than a
student who graduates from college B.

10.3 Comparing Two Independent Population Means with Known Population Standard Deviations[3]

Even though this situation is not likely (knowing the population standard deviations is not likely), the
following example illustrates hypothesis testing for independent means, known population standard deviations. The distribution is Normal and is for the difference of sample means, $\overline{X_1} - \overline{X_2}$. The normal
distribution has the following format:

Normal distribution

$$\overline{X_1} - \overline{X_2} \sim N\left[u_1 - u_2, \sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}\right] \tag{10.4}$$

The standard deviation is:

$$\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}} \tag{10.5}$$

The test statistic (z-score) is:

$$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}} \tag{10.6}$$

---

[3]This content is available online at <http://http://cnx.org/content/m17042/1.8/>.

Example 10.3

independent groups, population standard deviations known: The mean lasting time of 2 competing floor waxes is to be compared. **Twenty floors** are randomly assigned **to test each wax**. The following table is the result.

Wax	Sample Mean Number of Months Floor Wax Last	Population Standard Deviation
1	3	0.33
2	2.9	0.36

Table 10.2

Problem

Does the data indicate that **wax 1 is more effective than wax 2**? Test at a 5% level of significance.

Solution

This is a test of two independent groups, two population means, population standard deviations known.

Random Variable: $\overline{X_1} - \overline{X_2}$ = difference in the average number of months the competing floor waxes last.

$H_o : \mu_1 \leq \mu_2$

$H_a : \mu_1 > \mu_2$

The words **"is more effective"** says that **wax 1 lasts longer than wax 2**, on the average. "Longer" is a $">"$ symbol and goes into H_a. Therefore, this is a right-tailed test.

Distribution for the test: The population standard deviations are known so the distribution is normal. Using the formula above, the distribution is:

$$\overline{X_1} - \overline{X_2} \sim N\left(0, \sqrt{\frac{0.33^2}{20} + \frac{0.36^2}{20}}\right)$$

Since $\mu_1 \leq \mu_2$ then $\mu_1 - \mu_2 \leq 0$ and the mean for the normal distribution is 0.

Calculate the p-value using the normal distribution: p-value = 0.1799

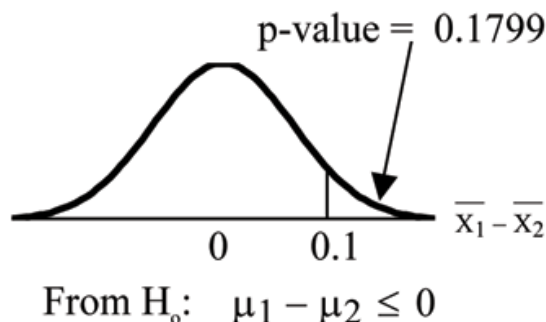**Graph:**

$$\text{p-value} = 0.1799$$

From $H_o:$ $\mu_1 - \mu_2 \leq 0$

Figure 10.2

$\overline{x_1} - \overline{x_2} = 3 - 2.9 = 0.1$

Compare α and the p-value: $\alpha = 0.05$ and p-value = 0.1799. Therefore, $\alpha <$ p-value.

Make a decision: Since $\alpha <$ p-value, do not reject H_o.

Conclusion: At the 5% level of significance, from the sample data, there is not sufficient evidence to conclude that wax 1 lasts longer (wax 1 is more effective) than wax 2.

NOTE: TI-83+ and TI-84: Press STAT. Arrow over to TESTS and press 3:2-SampZTest. Arrow over to Stats and press ENTER. Arrow down and enter .33 for sigma1, .36 for sigma2, 3 for the first sample mean, 20 for n1, 2.9 for the second sample mean, and 20 for n2. Arrow down to μ1: and arrow to $> \mu$2. Press ENTER. Arrow down to Calculate and press ENTER. The p-value is p = 0.1799 and the test statistic is 0.9157. Do the procedure again but instead of Calculate do Draw.

10.4 Comparing Two Independent Population Proportions[4]

1. The two independent samples are simple random samples that are independent.
2. The number of successes is at least five and the number of failures is at least five for each of the samples.

Comparing two proportions, like comparing two means, is common. If two estimated proportions are different, it may be due to a difference in the populations or it may be due to chance. A hypothesis test can help determine if a difference in the estimated proportions $(P'_A - P'_B)$ reflects a difference in the populations.

The difference of two proportions follows an approximate normal distribution. Generally, the null hypothesis states that the two proportions are the same. That is, $H_0 : p_A = p_B$. To conduct the test, we use a pooled proportion, p_c.

[4]This content is available online at <http://http://cnx.org/content/m17043/1.10/>.

The pooled proportion is calculated as follows:

$$p_c = \frac{X_A + X_B}{n_A + n_B} \qquad (10.7)$$

The distribution for the differences is:

$$P'_A - P'_B \sim N\left[0, \sqrt{p_c \cdot (1 - p_c) \cdot \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}\right] \qquad (10.8)$$

The test statistic (z-score) is:

$$z = \frac{(p'_A - p'_B) - (p_A - p_B)}{\sqrt{p_c \cdot (1 - p_c) \cdot \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \qquad (10.9)$$

Example 10.4: Two population proportions

Two types of medication for hives are being tested to determine if there is a **difference in the percentage of adult patient reactions. Twenty** out of a random **sample of 200** adults given medication A still had hives 30 minutes after taking the medication. **Twelve** out of another **random sample of 200 adults** given medication B still had hives 30 minutes after taking the medication. Test at a 1% level of significance.

10.4.1 Determining the solution

This is a test of 2 population proportions.

Problem   *(Solution on p. 460.)*
  How do you know?

Let A and B be the subscripts for medication A and medication B. Then p_A and p_B are the desired population proportions.

Random Variable:
$P'_A - P'_B$ = difference in the percentages of adult patients who did not react after 30 minutes to medication A and medication B.

$H_o : p_A = p_B \qquad\qquad p_A - p_B = 0$

$H_a : p_A \neq p_B \qquad\qquad p_A - p_B \neq 0$

The words **"is a difference"** tell you the test is two-tailed.

Distribution for the test: Since this is a test of two binomial population proportions, the distribution is normal:

$p_c = \frac{X_A + X_B}{n_A + n_B} = \frac{20 + 12}{200 + 200} = 0.08 \quad 1 - p_c = 0.92$

Therefore, $P'_A - P'_B \sim N\left[0, \sqrt{(0.08) \cdot (0.92) \cdot \left(\frac{1}{200} + \frac{1}{200}\right)}\right]$

$P'_A - P'_B$ follows an approximate normal distribution.

Calculate the p-value using the normal distribution: p-value = 0.1404.

Estimated proportion for group A: $\quad p'_A = \frac{X_A}{n_A} = \frac{20}{200} = 0.1$

Estimated proportion for group B: $\quad p'_B = \frac{X_B}{n_B} = \frac{12}{200} = 0.06$
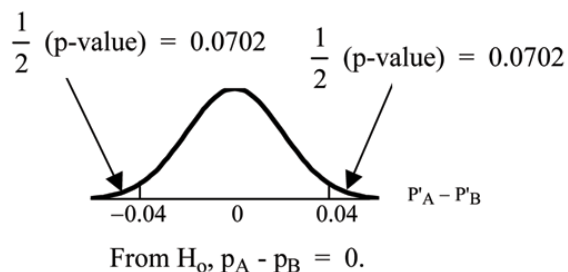
Graph:

Figure 10.3

$P'_A - P'_B = 0.1 - 0.06 = 0.04.$

Half the p-value is below -0.04 and half is above 0.04.

Compare α and the p-value: $\alpha = 0.01$ and the p-value $= 0.1404$. $\alpha < $ p-value.

Make a decision: Since $\alpha < $ p-value, you cannot reject H_o.

Conclusion: At a 1% level of significance, from the sample data, there is not sufficient evidence to conclude that there is a difference in the percentages of adult patients who did not react after 30 minutes to medication A and medication B.

NOTE: TI-83+ and TI-84: Press STAT. Arrow over to TESTS and press 6:2-PropZTest. Arrow down and enter 20 for $x1$, 200 for $n1$, 12 for $x2$, and 200 for $n2$. Arrow down to p1: and arrow to not equal p2. Press ENTER. Arrow down to Calculate and press ENTER. The p-value is $p = 0.1404$ and the test statistic is 1.47. Do the procedure again but instead of Calculate do Draw.

10.5 Matched or Paired Samples[5]

1. Simple random sampling is used.
2. Sample sizes are often small.
3. Two measurements (samples) are drawn from the same pair of individuals or objects.
4. Differences are calculated from the matched or paired samples.
5. The differences form the sample that is used for the hypothesis test.

---

[5]This content is available online at <http://http://cnx.org/content/m17033/1.12/>.

6. The matched pairs have differences that either come from a population that is normal or the number of differences is greater than 30 or both.

In a hypothesis test for matched or paired samples, subjects are matched in pairs and differences are calculated. The differences are the data. The population mean for the differences, μ_d, is then tested using a Student-t test for a single population mean with $n - 1$ degrees of freedom where n is the number of differences.

The test statistic (t-score) is:

$$t = \frac{\overline{x_d} - \mu_d}{\left(\frac{s_d}{\sqrt{n}}\right)} \tag{10.10}$$

Example 10.5: Matched or paired samples
A study was conducted to investigate the effectiveness of hypnotism in reducing pain. Results for randomly selected subjects are shown in the table. The "before" value is matched to an "after" value.

Subject:	A	B	C	D	E	F	G	H
Before	6.6	6.5	9.0	10.3	11.3	8.1	6.3	11.6
After	6.8	2.4	7.4	8.5	8.1	6.1	3.4	2.0

Table 10.3

Problem
Are the sensory measurements, on average, lower after hypnotism? Test at a 5% significance level.

Solution
Corresponding "before" and "after" values form matched pairs.

After Data	Before Data	Difference
6.8	6.6	0.2
2.4	6.5	-4.1
7.4	9	-1.6
8.5	10.3	-1.8
8.1	11.3	-3.2
6.1	8.1	-2
3.4	6.3	-2.9
2	11.6	-9.6

Table 10.4

The data **for the test** are the differences: {0.2, -4.1, -1.6, -1.8, -3.2, -2, -2.9, -9.6}

The sample mean and sample standard deviation of the differences are: $\overline{x_d} = -3.13$ and $s_d = 2.91$ Verify these values.

Let μ_d be the population mean for the differences. We use the subscript d to denote "differences."

Random Variable: $\overline{X_d}$ = the average difference of the sensory measurements

$$H_o : \mu_d \geq 0 \tag{10.11}$$

There is no improvement. (μ_d is the population mean of the differences.)

$$H_a : \mu_d < 0 \tag{10.12}$$

There is improvement. The score should be lower after hypnotism so the difference ought to be negative to indicate improvement.

Distribution for the test: The distribution is a student-t with $df = n - 1 = 8 - 1 = 7$. Use t_7. **(Notice that the test is for a single population mean.)**

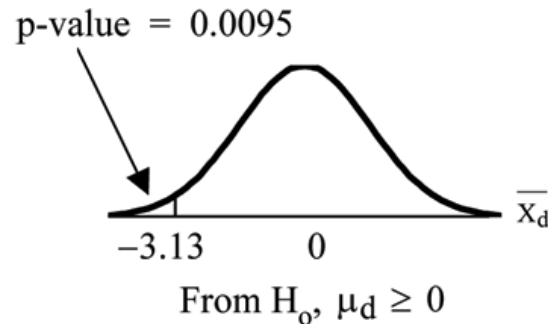**Calculate the p-value using the Student-t distribution:** p-value = 0.0095

Graph:

Figure 10.4

$\overline{X_d}$ is the random variable for the differences.

The sample mean and sample standard deviation of the differences are:

$\overline{x_d} = -3.13$

$\overline{s_d} = 2.91$

Compare α and the p-value: $\alpha = 0.05$ and p-value = 0.0095. $\alpha > $ p-value.

Make a decision: Since $\alpha > $ p-value, reject H_o.

This means that $\mu_d < 0$ and there is improvement.

Conclusion: At a 5% level of significance, from the sample data, there is sufficient evidence to conclude that the sensory measurements, on average, are lower after hypnotism. Hypnotism appears to be effective in reducing pain.

NOTE: For the TI-83+ and TI-84 calculators, you can either calculate the differences ahead of time (**after - before**) and put the differences into a list or you can put the **after** data into a first list and

the **before** data into a second list. Then go to a third list and arrow up to the name. Enter 1st list name - 2nd list name. The calculator will do the subtraction and you will have the differences in the third list.

NOTE: TI-83+ and TI-84: Use your list of differences as the data. Press STAT and arrow over to TESTS. Press 2:T-Test. Arrow over to Data and press ENTER. Arrow down and enter 0 for μ_0, the name of the list where you put the data, and 1 for Freq:. Arrow down to μ: and arrow over to $<$ μ_0. Press ENTER. Arrow down to Calculate and press ENTER. The p-value is 0.0094 and the test statistic is -3.04. Do these instructions again except arrow to Draw (instead of Calculate). Press ENTER.

Example 10.6
A college football coach was interested in whether the college's strength development class increased his players' maximum lift (in pounds) on the bench press exercise. He asked 4 of his players to participate in a study. The amount of weight they could each lift was recorded before they took the strength development class. After completing the class, the amount of weight they could each lift was again measured. The data are as follows:

Weight (in pounds)	Player 1	Player 2	Player 3	Player 4
Amount of weighted lifted prior to the class	205	241	338	368
Amount of weight lifted after the class	295	252	330	360

Table 10.5

The coach wants to know if the strength development class makes his players stronger, on average.

Problem   *(Solution on p. 460.)*
Record the **differences** data. Calculate the differences by subtracting the amount of weight lifted prior to the class from the weight lifted after completing the class. The data for the differences are: {90, 11, -8, -8}

Using the differences data, calculate the sample mean and the sample standard deviation.

$\bar{x}_d = 21.3$ $s_d = 46.7$

Using the difference data, this becomes a test of a single _____ (fill in the blank).

Define the random variable: $\bar{X}_d$ = average difference in the maximum lift per player.

The distribution for the hypothesis test is t_3.

$H_o : \mu d \leq 0$ $H_a : \mu_d > 0$

Graph:

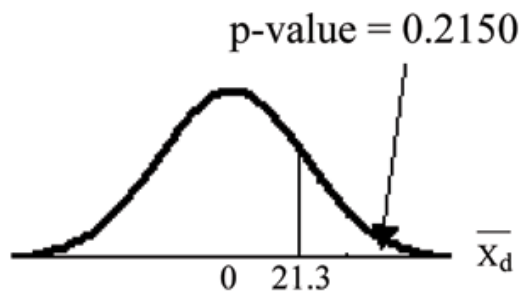p-value = 0.2150

0   21.3 $\overline{X_d}$

Figure 10.5

Calculate the p-value: The p-value is 0.2150

Decision: If the level of significance is 5%, the decision is to not reject the null hypothesis because $\alpha < $ p-value.

What is the conclusion?

Example 10.7
 Seven eighth graders at Kennedy Middle School measured how far they could push the shot-put with their dominant (writing) hand and their weaker (non-writing) hand. They thought that they could push equal distances with either hand. The following data was collected.

Distance (in feet) using	Student 1	Student 2	Student 3	Student 4	Student 5	Student 6	Student 7
Dominant Hand	30	26	34	17	19	26	20
Weaker Hand	28	14	27	18	17	26	16

Table 10.6

Problem *(Solution on p. 460.)*
Conduct a hypothesis test to determine whether the differences in distances between the children's dominant versus weaker hands is significant.

HINT: use a t-test on the difference data.

CHECK: The test statistic is 2.18 and the p-value is 0.0716.

What is your conclusion?

10.6 Summary of Types of Hypothesis Tests[6]

Two Population Means

- Populations are independent and population standard deviations are unknown.
- Populations are independent and population standard deviations are known (not likely).

Matched or Paired Samples

- Two samples are drawn from the same set of objects.
- Samples are dependent.

Two Population Proportions

- Populations are independent.

---

[6]This content is available online at <http://http://cnx.org/content/m17044/1.5/>.